MLC 2019 submission description

Çağrı Çöltekin

ccoltekin@sfs.uni-tuebingen.de

Taraka Rama

tarakark@ifi.uio.no

July 1, 2019

This submission introduces one new measure, *inflection accuracy*, and also includes recalculation of our previous measures from MLC 2018 on the current set of UD treebanks. In this document we briefly describe inflection accuracy, and refer to Çöltekin and Rama (2018) for the description of the other measures.

Inflection accuracy is simply the accuracy of a machine learning model predicting the inflected word from its lemma and morphological features. The intuition is that if the language in question has a complex morphology, the accuracy is expected to be low. Our target was mainly to detect 'non-transparent', irregular morphology, but initial results does not fully agree with the expectations (although they seem to put languages with similar morphological typology close to each other).

The morphological inflection systems has been popularized by successive SIG-MORPHON shared tasks (Cotterell et al. 2017; Cotterell et al. 2018). The inflection system used for this measure is described in Çöltekin (2019). The method is based on linear classifiers, which seem to provide close to state-of-the-art systems with relatively small demand on computing power. Our focus here is on the differences in inflection accuracy rather than the overall success of the system in the inflection task.

The results reported here are obtained using a random sample of 1000 words. The reported scores are the mean accuracy (exact match of the inflected word) over 3-fold cross validation on this sample. The model is tuned for each language separately using a random search through the model parameters.

Similar to the MLC 2018 shared task, the earlier measures were calculated using sampling. For this set of results we use a sample size of 20 000 tokens, and report the average of 1000 samples.

The code producing both old and the new measures are available at https://github.com/coltekin/mlc2018. The README file on the repository describes the usage and requirements of the programs.

We have also experimented with a few other methods (including some experiments on the PBC data). However, we do not have stable results at this time.

References

Çöltekin, Çağrı (2019). "Cross-lingual morphological inflection with explicit alignment." In: *Proceedings of the 16th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, (to appear).

Çöltekin, Çağrı and Taraka Rama (2018). "Exploiting Universal Dependencies Treebanks for Measuring Morphosyntactic Complexity." In: *Proceedings of First Workshop on Measuring Language Complexity*. Ed. by Christian Bentz and Aleksandrs Berdicevskis. Torun, Poland, pp. 1–7.

Cotterell, Ryan, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Arya D. McCarthy, Katharina Kann, Sebastian Mielke, Garrett Nicolai, Miikka Silfverberg, David Yarowsky, Jason Eisner, and Mans Hulden (Oct. 2018). "The CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection." In: *Proceedings of the CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*. Brussels: Association for Computational Linguistics, pp. 1–27. URL: https://www.aclweb.org/anthology/K18-3001.

Cotterell, Ryan, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sandra Kübler, David Yarowsky, Jason Eisner, and Mans Hulden (Aug. 2017). "CoNLL-SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection in 52 Languages." In: *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*. Vancouver: Association for Computational Linguistics, pp. 1–30. DOI: 10.18653/v1/K17-2001. URL: https://www.aclweb.org/anthology/K17-2001.